Journal of Heredity 2010:101(6):757–768
doi:10.1093/jhered/esq078
Advance Access publication June 27, 2010

Genetic Analysis of Genes Controlling Natural Variation of Seed Coat and Flower Colors in Soybean

Kiwoung Yang*, Namhee Jeong*, Jung-Kyung Moon, Yeong-Ho Lee, Suk-Ha Lee, Hwan Mook Kim, Cheol Ho Hwang, Kyoungwhan Back, Reid G. Palmer, and Soon-Chun Jeong

From Bio-Evaluation Center, Korea Research Institute of Bioscience and Biotechnology, Cheongwon, Chungbuk 363-883, Republic of Korea (Yang, Jeong, Kim, S.-C. Jeong); the National Institute of Crop Science, Rural Development Administration, Suwon, Republic of Korea (Moon, Y.-H. Lee); the Department of Plant Science and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea (S.-H. Lee); the Department of Crop Science and Biotechnology, Dankook University, Chungnam, Republic of Korea (Hwang); the Molecular Biotechnology Major, Chonnam National University, Gwangju, Republic of Korea (Back); and the USDA ARS Corn Insects and Crop Genetics Research Unit, Department of Agronomy, Iowa State University, Ames, IA (Palmer).

*These authors contributed equally to the work.

Address correspondence to Soon-Chun Jeong at the address above, or e-mail: scjeong@kribb.re.kr.

Abstract

Soybean exhibits natural variation in flower and seed coat colors via the deposition of various anthocyanin pigments in the respective tissues. Although pigmentation in seeds or flowers has been well dissected at molecular level in several plant species, the genes controlling natural variation in anthocyanin traits in the soybean are not completely understood. To evaluate the genetic correlation between genetic loci and genes, 8 enzyme-encoding gene families and a transcription factor were localized in a soybean genome-wide genetic map. Among the seed coat color–controlling loci, the genetic location of the gene encoding for *W1* was substantiated in the context of the current soybean molecular genetic map and *O* was postulated to correspond to anthocyanidin reductase. Among the genetic loci that regulate flower pigmentation, the genetic locations of the genes encoding for *W1*, *W4*, and *Wp* were identified, *W3* was mapped on soybean linkage group B2 (chromosome 14), and *W2* was postulated to correspond to an MYB transcription factor. Correlation studies between the developed markers and 3 color-controlling loci provided important empirical data that should prove useful in the design of marker-assisted breeding schemes as well as future association studies involving soybean.

Key words: *anthocyanin, flower color,* Glycine max, *molecular marker, seed coat color, soybean*

Introduction

Colors are one of the earliest characteristics assessed in genetic studies, since the work conducted by Mendel. Genetic studies conducted on the soybean [*Glycine max* (L.) Merr.] in the early 1900s following the rediscovery of Mendel's laws were concerned primarily with color characteristics (Woodworth 1921). Although pigmentation in seeds or flowers has been extensively studied in maize, *Antirrhinum*, petunia, *Arabidopsis*, etc. (Lepiniec et al. 2006), it remains poorly understood at the molecular genetic level in soybean. This is mainly because the flower, seed coat, and pubescence colors of soybean are of limited utility, except for their use as markers to assess hybridity in breeding programs. Several recent studies have concerned partial seed coat pigmentation as the result of chilling stress or viral diseases as the pigmentation degrades the external appearance of soybean seeds and reduces their commercial value (Takahashi and Asanuma 1996; Takahashi 1997; Gore et al. 2002; Benitez et al. 2004; Senda et al. 2004). However, natural products that cause colors, including flavonoids and anthocyanins, are currently attracting attention because of their medicinal and nutritional values, which are the result of their antioxidant properties and flavors (reviewed in Dixon and Sumner 2003).

Pigmentation of the flowers and seed coats is induced via the deposition of a number of flavonoids in the respective tissues of the soybean. The synthesis of these compounds mainly derives from an anthocyanin biosynthesis branch of

the phenylpropanoid pathway of secondary metabolism. Thus far, alleles of at least 5 genetic loci (*I*, *T*, *W1*, *R*, and *O*) are known to function epistatically to control seed coat pigmentation and 6 genetic loci (*W1*, *W2*, *W3*, *W4*, *Wm*, and *Wp*) to control flower pigmentation (reviewed by Palmer et al. 2004). Three independent loci (*I*, *R*, and *T*) mainly control the biosynthesis of the pigments that determine the seed coat colors. The *O* and *W1* loci influence seed coat color only in the homozygous recessive *ir* or *it* genotypes, respectively (Palmer et al. 2004). The *I* locus is located at a region harboring a cluster of chalcone synthase (CHS) genes on soybean molecular linkage group (MLG) A2, which is chromosome 8 (Todd and Vodkin 1996; Clough et al. 2004; Tuteja et al. 2004) and controls the distribution of anthocyanin and proanthocyanidin pigments. The dominant *I* allele exhibits a completely colorless seed coat phenotype due to the dominance inhibition via a possible posttranscriptional mode of gene silencing (Senda et al. 2004). The homozygous recessive *i* allele gives rise to self-colored seed coat. The other homozygous recessive alleles *i-i* and *i-k* give rise to the restriction of pigment distribution to specific regions of the seed coat. The *R* and *T* loci control specific seed coat colors, which include black (*i*, *R*, *T*), imperfect black (*i*, *R*, *t*), brown (*i*, *r*, *T*), or buff (*i*, *r*, *t*), by controlling the types of the anthocyanin and proanthocyanidin pigments. The principal anthocyanin pigments in the seed coats of black soybeans are cyanidin-3-monoglucoside and delphinidin-3-monoglucoside. Pelargonidin-3-glucoside is not a major anthocyanin and is generated in special cases (Lee et al. 2009). Under the presence of the dominant *I* allele, 3 independent genes *k1*, *k2*, and *k3* generate dark or saddle colors on seed coat (Palmer et al. 2004). However, their relationship to the *T, R,* and *W1* loci has yet to be completed (Rode and Bernard 1975). The cloning and mapping of the soybean F3′H genomic and cDNA sequences showed that the *F3′H* gene cosegregates with the *T* locus on soybean linkage group C2 (chromosome 6) (Toda et al. 2002; Zabala and Vodkin 2003). The *R* locus initially was mapped to MLG K (chromosome 9) in a population generated from a cross between Minsoy and Noir 1 (Lark et al. 1993) and was later located between A668_1 and K387_1 (Song et al. 2004). However, molecular nature of the *R* locus has yet to be determined. The majority of cultivars exhibit either purple (*W1*) or white (*w1*) flowers. Chromatographic experiments and genetic analysis suggested that *W1* encodes a flavonoid 3' 5' hydroxylase (F3′5′H) (Buzzell et al. 1987; Zabala and Vodkin 2007). The homozygous genotype *W1w2* generates purple–blue flower. The genotype *W1W3w4* generates the purple-throat phenotype, and the genotype *W1w3w4* generates the near-white phenotype. The *W3* locus cosegregates with dihydroflavonol 4-reductase (DFR) on the bases of restriction fragment length polymorphism (RFLP) analysis of isolines and an F_2 population (Fasoula et al. 1995). Magenta and pink flower colors result from the *W1wm* and from the *W1wp* genotypes, respectively (Palmer and Kilen 1987; Stephens and Nickell 1992). The *Wm* locus is associated with flavonol synthase (FLS) that is not required for production of anthocyanin pigments but may compete with the products of F3′5′H and DFR for common dihydroflavonol substrates (Takahashi et al. 2007). The *Wp* locus encodes the flavonoid 3-hydroxylase 1 gene on the bases of a series of genetic analyses (Zabala and Vodkin 2005).

Molecular markers developed from the genes predicted for the *I*, *T*, and *Wm* loci cosegregate with the presumed genetic loci in mapping populations and were placed on the soybean molecular linkage maps. However, in the case of the *W1*, *W3*, *W4*, and *Wp* loci, the markers developed from the predicted genes were shown to cosegregate with the genetic loci in isolines or mapping populations but were not substantiated in the context of the current soybean molecular linkage map. The genetic map locations of the *W2*, *O*, and *R* loci were known by molecular or classical genetic mappings, but the molecular nature of genes encoding these loci have yet to be determined. The different flower and seed coat colors may be reflective of mutations that affect enzymes at different steps of the anthocyanin biosynthetic pathway. In this study, we have developed sequence-based markers using sequence information from the genes involved in the anthocyanin metabolic pathway. This work has resulted in the formulation of several specific genetic hypotheses that suggest correlations between classical genetic loci and the genes involved in anthocyanin metabolism. In addition, correlation studies between molecular markers and color-regulating loci, whose genotypes could be visually evaluated, provide a guideline for future studies of soybean genome-wide association.

Materials and Methods

Plant Materials

A population (HI) was generated by an interspecific cross between a *G. max* line "Hwangkeum" and a *G. soja* (Siebold & Zucc.) line "IT182932." In total, 113 F_2-derived F_{12} recombinant inbred lines (RILs) were utilized in the construction of a framework map consisting of 20 soybean linkage groups in this study. The genome-wide genetic map was constructed using 421 markers of the HI population (Yang et al. 2008). The map was utilized to determine genetic map locations of anthocyanin-related genes. The genetic correlation between markers and colors, which are expressed by the *I*, *T*, and *W1* loci, were assessed in a collection of 18 presumably diverse soybean lines developed or collected in the United States and Korea: Hwangkeum, Pureun, Sinpaldal, Sinpaldal 2, Sowon, SS2-2, and Taekwang developed in Korea; IT182932 collected in Korea; Evans, Hutcheson, L29, Lee68, Marshall, Peking, PI96983, V94-5152, Williams, and York developed or collected in the United States. An RIL population from the cross between *G. max* Pureun (white flower) and *G. max* Jinpum 2 (purple flower) that contains more than 300 mapped markers (Cai et al. 2008) was employed in order to confirm the cosegregation between the developed markers and the *W1* locus.

Development of Sequence-Based Markers from Genes in Anthocyanin Metabolism

The genomic or cDNA sequences of genes that previously were reported to be involved in the anthocyanin biosynthesis

pathway were utilized to develop markers to investigate the possible correspondences in genetic locations between the genes and the genetic loci involved in the regulation of seed coat and flower color. The soybean genes from which markers were developed included those encoding most of the enzymes of the anthocyanin biosynthetic pathway; inverted *CHS* repeats, flavonoid 3'-hydroxylase (*F3'H*), flavonoid 3' 5'-hydroxylase (*F3'5'H*), flavanone 3-hydroxylase (*F3H*), *DFR*, anthocyanidin synthase (*ANS*), anthocyanin reductase (*BANYULS*), and several uridine diphosphate-flavonoid glucosyltransferase (*UFGT*). Among several known regulatory genes including bHLH, MYB, and WD40-repeat transcription factors, which have been identified to be involved in the anthocyanin biosynthesis in plant species other than soybean (Lepiniec et al. 2006), the tomato *ANT1* gene (Mathews et al. 2003), which encodes for an R2R3 MYB domain–containing transcription factor, hit 2 soybean expressed sequence tags (ESTs), BM093788 and BM092559, highly homologous to the *ANT1* sequence in the conserved motif as well as the nonconserved region in a basic alignment search tool (BLAST) search against GenBank databases (November 2005). Soybean sequences that had been characterized as regulating anthocyanin biosynthesis, or were homologous to the genes characterized in other plants, first were used in the development of appropriate markers (Supplementary Table 1). In cases in which these initial attempts proved unsuccessful, their corresponding sequences in the soybean parental lines of our mapping populations were determined in order to locate polymorphic sites (the sequences deposited into GenBank were listed in Supplementary Table 1). Sequence analysis and marker development procedures were conducted as described (Jeong and Saghai Maroof 2004; Yang et al. 2008). Sets of allele-specific polymerase chain reaction (PCR) primers for the genotyping of single nucleotide polymorphisms (SNPs) or primer sets for the genotyping of sequence length polymorphisms were designed using Primer3 software (Rozen and Skaletsky 2000).

Analysis of Seed Coat and Flower Colors and Markers

Seed coat colors were assessed by inspecting mature and dried seeds visually. Flower colors were assessed visually by inspecting fully open flower petals. Young trifoliolate leaf tissues were collected from field or greenhouse-grown soybean plants. DNA was isolated via a cetyltrimethylammonium bromide extraction method (Doyle JJ and Doyle JV 1987). Primers for marker analysis were custom-made by Bioneer (Daejeon, Korea). Allele-specific SNP and simple sequence length polymorphism markers were analyzed as previously described (Jeong and Saghai Maroof 2004).

Linkage Analysis

Genotypes obtained from the developed markers and seed coat colors were combined with the genotypes of markers preexisting in the HI population in order to estimate their genetic linkage relationships. MapMaker 3.0b software was used for the grouping and ordering of the genetic loci (Lander et al. 1987). The mapping criteria and parameter settings, including a logarithm of odds of 5.0 and a maximum genetic distance of 37.5 cM, were identical to those described by Yang et al. (2008). The MLGs were named in accordance with the designations of the integrated soybean genetic map previously generated by Song et al. (2004). For the assignment of soybean molecular groups to chromosomes, see http://soybase.org/LG2Xsome.php.

Results

Mapping Strategy

A schematic diagram of the stages of the anthocyanin biosynthetic and metabolic pathway is provided in Figure 1. Markers were developed from the sequences of the genes encoding for the enzymes, as well as a regulator in the phenylpropanoid pathway that metabolizes anthocyanins. The genetic map locations for 8 enzyme-encoding gene families and a transcription factor from which markers were developed are illustrated in Table 1 (see Supplementary Table 1 and Figure 1 for detailed information). Hwangkeum and IT182932 showed a number of phenotypic differences, including seed coat color, seed weight, leaf size, soybean mosaic virus resistance, powdery mildew resistance, growth habit, number of branches, and pod shattering. In particular, diverse seed coat colors of the RILs in the HI population permitted us to infer approximate genetic locations of the loci controlling seed coat color variation (Figure 2; Supplementary Figure 1). For the *I*, *T*, and *W1* loci, whose genotypes in soybean accessions could be easily evaluated, correlation between molecular markers and color phenotypes was investigated in order to determine the utility of the developed markers as functional markers as well as to suggest a marker development strategy for future soybean genome-wide association studies.

Inverted CHS Repeats Cosegregate with the *I* Locus

Mutations of a dominant allele (*I* or *i-i*) to a recessive allele (*i*) were linked to structural changes taking place at an inverted repeat cluster of 6 *CHS* genes comprising the *I* locus encompassing a region of at least 27 kb (Tuteja et al. 2004). We presumed that it would be difficult to design primers from the repeated CHS genic regions to generate specific amplicons. Thus, primers were designed to PCR-amplify the microsatellite regions flanking the repeated-gene cluster, which was observed on the 103-kb bacterial artificial chromosome (BAC) clone (BAC104J7) sequence harboring the inverted perfect repeat cluster comprising the *I* locus (Clough et al. 2004). Four microsatellite markers, SM303, SM304, SM305, and SM306, were generated (Table 1; Supplementary Table 1). The genotypes of RIL individuals in the HI population determined by these markers corresponded completely with the genotypes of RIL individuals determined by colorless (yellow and green) (dominant, *I*) or any other colored genotypes (recessive, *i*). As expected, the *CHS* repeats and *I* locus were colocalized between A454.p2 and GMNOD2B on MLG A2 (chromosome 8) (Table 1).

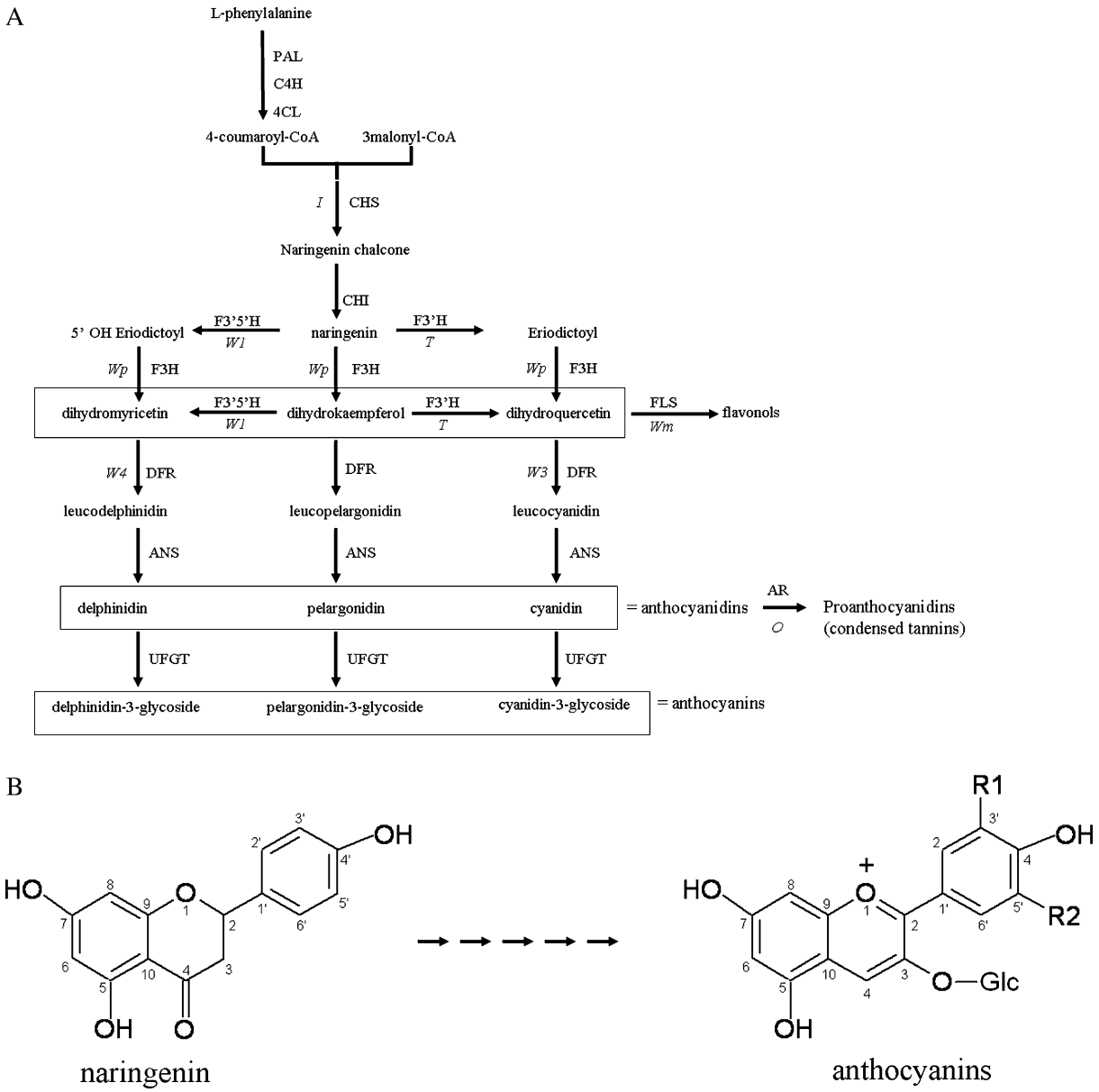

Figure 1. Scheme of the anthocyanin metabolic pathway (**A**) and chemical structures of naringenin and anthocyanin (**B**) in soybean. Enzymes and genetic loci (*I*, *R*, *T*, *W1*, *W3*, *Wm*, and *Wp*) associated with the indicated enzymes in the pathway are indicated along the arrows. ANS (also called LDOX, leucoanthocyanidin dioxygenase); AR, anthocyanidin reductase; C4H, cinnamic acid 4-hydroxylase; CHI, chalcone isomerase; 4CL, 4-coumarate: CoA ligase; CHS; DFR; F3H, flavanone 3-hydroxylase; F3′H, flavonoid 3'-hydroxylase; F3′5′H, flavonoid 3'5'-hydroxylase; FLS; PAL, phenylalanine ammonia-lyase; UFGT. Delphinidin-3-glucoside (R1=OH, R2=OH); pelargonidin-3-glucoside (R1=H, R2=H); cyanidin-3-glucoside (R1=OH, R2=H).

In order to evaluate the utility of the developed markers in a marker-assisted breeding program, marker-based genotypes were correlated with seed coat color–based genotypes. The SM303, SM304, SM305, and SM306 markers were scored in 16 diverse soybean lines via polyacrylamide gel separations (Figure 3). The size of the PCR bands was highly variable among the soybean accessions. Four different DNA fragments (alleles) of SM303, 2 alleles of SM304, 2 alleles of SM305, and 5 alleles of SM306 were detected. IT182932 and Peking, which have black seed coats expressed by the *i* allele, harbored the common smallest fragment (allele) at the SM303 and SM304 loci, but did contain different alleles at SM306. The second smallest fragment of SM303 was correlated with the *i-i* phenotype (yellow seed coat and colored hilum). The third allele of SM303, which had a slightly larger fragment than the second allele, was unique to PI96983. The PI96983 accession analyzed in this study showed light brown–colored hila but was listed as yellow-colored hilum in

Table 1 Postulated correlations between mapped genes and classical genetic loci controlling soybean seed coat and flower pigmentation

Gene name	Locus name	Marker name	Linkage group (chromosome number)	North marker (cM)[a]	South marker (cM)[a]
AR (anthocyanidin reductase)	*O*	SN013	A2 (8)	Satt207 (4.1)	Satt493 (0.0)
Inverted CHS repeats	*I*	SM304, SM305, SM303, SM306	A2 (8)	A454.p2 (11.0)	GMNOD2B (8.2)
ANS1	—	SL015, SN014	B1 (11)	BE806308 (10.6)	Sat_272 (3.1)
ANS2	—	SN015	D1a (1)	Sat_414 (11.4)	Satt129 (2.5)
DFR1	*W3*	SN016, SM004	B2 (14)	Sat_287 (5.4)	Satt467 (1.2)
DFR3	—	SM005, SM325	D1b (2)	Satt579 (0.0)	Satt428 (0.0)
DFR2	*W4*	SL016	D2 (12)	Satt386 (2.2)	Sct_137 (2.8)
MYB transcription factor	*W2*	SL017	B2 (14)	Sat_009 (1.9)	Satt070 (1.9)
F3'H (flavonoid 3' hydroxylase)	*T*	SL305, SN317	C2 (6)	Satt286 (7.4)	Satt365 (0.5)
Anthocyanidin 3-O-glucosyltransferase	—	SM326, SM327	D1a (1)	AW781285 (3.6)	Sat_305 (1.1)
Anthocyanidin 3-O-glucosyltransferase	—	SM328	D1b (2)	CA783003S (4.4)	Satt558 (6.1)
Anthocyanidin 3-O-glucosyltransferase	—	SM329	F (13)	SOYHSP176 (0.0)	N11PFSCAR219 (3.3)
Anthocyanidin 3-O-glucosyltransferase	—	SN017	J (16)	Sctt011 (5.8)	Sat_396 (2.1)
Anthocyanidin 3-O-glucosyltransferase	—	SN018	J (16)	Sctt011 (7.5)	Sat_396 (0.4)
F3H1 (flavanone 3-hydroxylase)	*Wp*	SN318	D1b (2)	SL007 (2.0)	Satt216 (1.7)
F3H2 (flavanone 3-hydroxylase)	—	SN319	D1b (2)	SL007 (2.0)	Satt216 (1.7)
F3'5'H (flavonoid 3'5' hydroxylase)	*W1*	SN019, SL018 SN020, SL019	F (13)	Satt348 (0.0)	Satt160 (0.0)

[a] North and south markers (their centiMorgan [cM] distances) that are the most closely linked to markers developed from the anthocyanin-related genes on the soybean genome-wide genetic map reported by Yang et al. (2008).

Soybase (http://soybase.ncgr.org/cgi-bin/ace/generic/search/soybase as of July 2009). The fourth allele of SM303 was correlated with the *I* phenotype (yellow seed coat and yellow hilum color). Hutcheson (Buss et al. 1988) gave 2 alleles of SM303 and SM306 in repeated attempts that used genomic DNAs extracted from 2 individual plants belonging to the same Hutcheson seed lot, suggesting that the structure of the Hutcheson genome needs to be evaluated. Thus, the SM303-based genotypes are perfectly correlated with seed coat color–based genotypes. The SM304 marker did not distinguish *i-i* from *I*. The SM305 marker distinguished between *I* and recessive *i* alleles. The size of the DNA fragment of SM306 proved highly variable in the set of 16 soybean lines, but no correlations were discovered between the band size and the *I* locus type. SM306 did not distinguish among *i*, *i-i*, and *I*. For example, IT182932, Taekwang, Lee68, which contain *i*, *I*, and *i-i*, respectively, contain the common smallest DNA fragment of SM306.

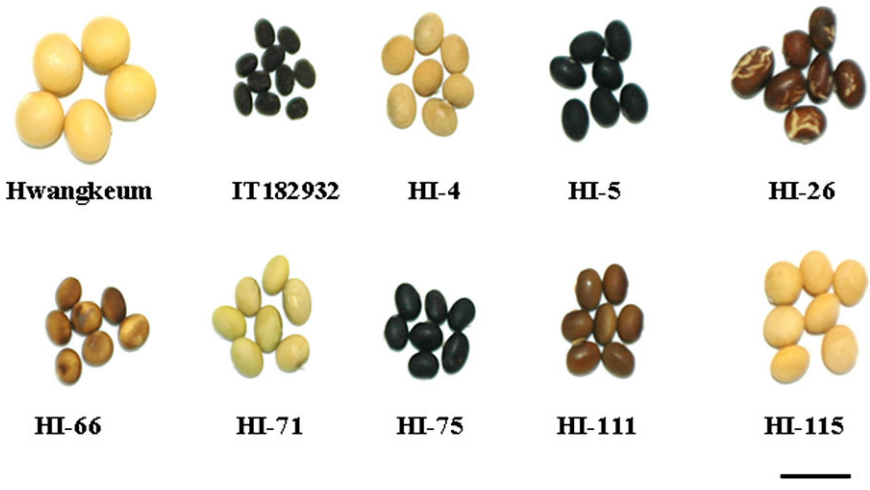

Figure 2. The polymorphism of soybean seed shape and coat color of the population generated via an interspecific cross between a *Glycine max* line Hwangkeum and a *G. soja* line IT182932. The RILs were designated by HI numbers. Scale bar, 10 mm.

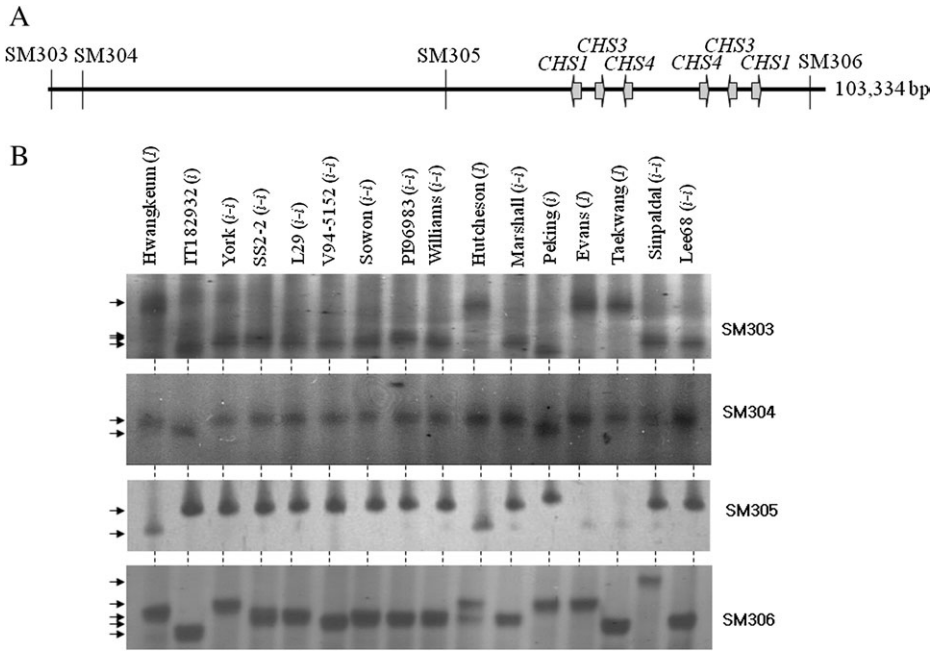

Figure 3. Schematic representation showing positions of the sequence-based SM303, SM304, SM305, and SM306 markers and positions and relative orientations of *CHS1*, *CHS3*, and *CHS4*, as marked by the block arrows, on BAC104J7 comprising the *I* locus (**A**) and polyacrylamide gels showing the separation of alleles of the sequence-based SM303, SM304, SM305, and SM306 markers in 16 diverse soybean accessions with various *I* locus genotypes (**B**). Distinct alleles detected in each gel are indicated by arrows. The lanes that contain the common soybean accession between gels are joined by lines.

Flavonoid 3′-Hydroxylase Cosegregating with *T*

The genomic sequences of the F3′H genome sequence that was identified as the *T* locus (Toda et al. 2002; Zabala and Vodkin 2003) were utilized to generate molecular markers. Two interesting insertion/deletion (indel) positions were observed from the alignment of the cDNA and genomic DNA sequences (GenBank accession no. AF501293–AF501305) of the F3′H genes. First, the predicted intron II sequences of the F3′H gene in XB22A (*T*) and Williams (*T*) harbor a 262 bp deletion relative to those observed in Richland(*t*), T157(*t*), and Harosoy (*t*) (Zabala and Vodkin 2003). The 262-bp indel site was utilized to generate the SL305 marker (Yang et al. 2008). Second, in Richland (*t*), T157 (*t*), and Harosoy (*t*), a base (C) deletion that was presumed to induces premature termination of the open reading frame (Zabala and Vodkin 2003) was detected at the coding region, 279 bp from the 3′ end of intron II. This presumed functional nucleotide polymorphism was utilized in the generation of an allele-specific PCR marker, SN317. The 2 markers were mapped between Satt286 and Satt365 on MLG C2 (chromosome 6), which was identified as the position of the *T* locus (Toda et al. 2002) (Table 1).

The diversity of the markers was investigated in 16 soybean lines (Figure 4). Only 2 alleles were observed for both markers. The plant genotypes scored by the 2 markers were identical. Considering the *I* allele–containing plant lines, in which the genotypes of the *T* locus could not be determined, the genotypes scored by the 2 molecular markers correlated perfectly with the genotypes scored by seed coat and hilum colors. In other words, all cultivars that could be predicted to contain the dominant *T* allele on the basis of seed coat color and hilum color also harbored the same allele as that of IT182932 (black seed coat and hilum, *T*) at the marker loci.

Flavonoid 3′ 5′-Hydroxylase Cosegregating with *W1*

The *F3′5′H* gene was identified by other researchers as being associated with the *W1* locus on the basis of analyses of an F3′5′H mutant, a series of soybean cultivars and lines, and an F_2 population (Zabala and Vodkin 2007). However, the genetic map location of the marker generated from the 65-bp insertion of the mutant allele (*w1*) relative to the *W1* allele of the *F3′5′H* gene was not substantiated by genetic linkage mapping. In this study, we also utilized the *F3′5′H* cDNA sequence with GenBank accession no. AY117551 (deposited by R-M. Liao, T-M. Chu, C-S. Wang in 2003), which was used by Zabala and Vodkin (2007), to amplify genomic DNAs from Hwangkeum, IT182932, Pureun, and Jinpum 2 via PCR. One genomic DNA sequence corresponded to portions of the first intron and its flanking exons and another corresponded to a portion of the third exon of the F3′5′H genomic DNA sequence reported by Zabala and Vodkin (2007). One SNP locus that resided within the predicted intron I was utilized for the development of the SN019 marker (Figure 5). Another SNP locus that resided within the predicted exon III was utilized to generate the

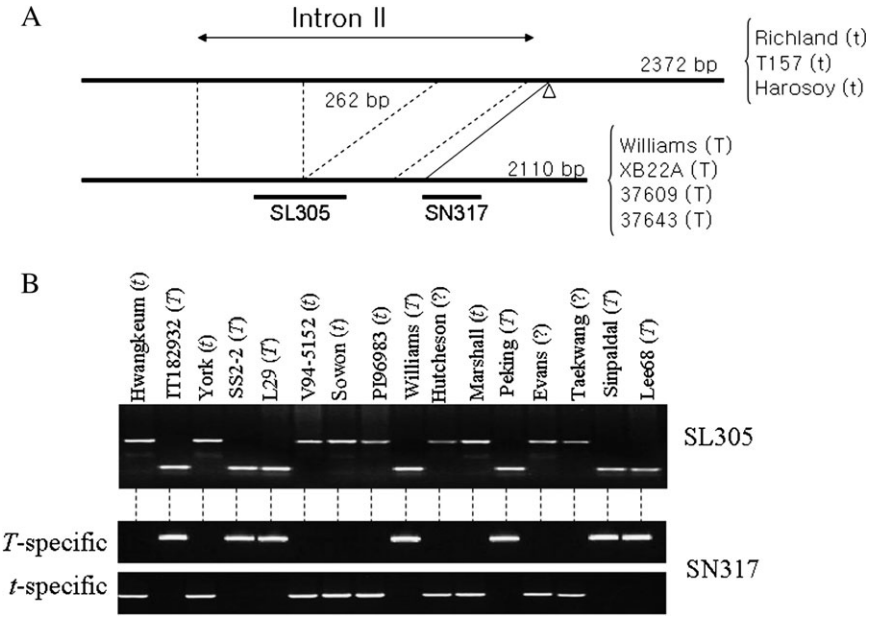

Figure 4. Schematic representation showing positions of the SL305 and SN317 markers on the genomic DNA segments encoding for *T* alleles (Zabala and Vodkin 2003) (**A**) and agarose gels showing the separation of the alleles of the sequence-based markers SL305 and SN317 in 16 diverse soybean accessions with *T* or *t* genotypes (**B**). The SN317 marker on the basis of the one base deletion of *t* allele was assayed using an allele-specific PCR procedure, as described by Jeong and Saghai Maroof (2004). The lanes that contain the common soybean accession between gels are joined by lines.

SN020 marker (Figure 5). The 2 markers cosegregated and were mapped between Satt348 and Satt160 on MLG F (chromosome 13), which is the presumed location of the *W1* locus in the public soybean map (Song et al. 2004). As both Hwangkeum and IT182932 have purple flowers, the Pureun (*w1*) × Jinpum 2 (*W1*) population (Cai et al. 2008) was employed in order to substantiate the cosegregation between the developed markers and the *W1* locus. The genotypes of the indel-based SL018 and SL019 markers that were generated from the predicted intron I and the predicted exon III, respectively, cosegregated with the genotypes determined by purple or white flowers in the Pureun × Jinpum 2 population. SL018, SL019, and *W1* that were mapped between Satt252 and Satt160 on MLG F (chromosome 13) were located 5.7 cM away from Satt252 and 0 cM away from Satt160 in the Pureun × Jinpum 2 population.

The correlation of the SN019, SL018, SN020, and SL019 markers with purple or white flower colors was assessed in 18 diverse soybean lines via agarose gel separations. The genotypes scored by the SL019 marker were correlated perfectly with the genotypes scored by flower colors. Allelic diversity of SL018 and SN020 was low and the rare alleles detected were confined to the *W1* genotypes. In the case of SL018, an additional 80 Korean soybean cultivars were assessed. The rare allele unique to Jinpum 2 was detected in 8 cultivars that contain purple flowers (data not shown). The Hwangkeum-specific allele of SN019 was detected in 7 soybean accessions that have the *W1* allele. However, the IT182932-specific allele of SN019 did not distinguish between *W1* and *w1*. Taken together, the diversity analyses indicated that the mutations at SL018, SN019, and SN020 occurred later than that at SL019, which detects the causative polymorphism proposed by Zabala and Vodkin (2007).

Flavonoid 3-Hydroxylase Cosegregating with *Wp*

The *Wp* locus was located on MLG D1b (chromosome 2) (Hegstad et al. 2000) and encodes for the flavonoid 3-hydroxylase 1 (*F3H1*) gene (Zabala and Vodkin 2005). However, the genetic linkage relationship between *Wp* and *F3H* with public molecular markers has yet to be determined. The genomic DNA sequences of the *F3H1* and *F3H2* genes were employed to generate the molecular markers SN318 and SN319, respectively. Both of these markers cosegregated within the HI population, thereby suggesting that the *F3H1* and *F3H2* genes are tandem duplicates in this genomic region. The 2 genes were located between Sat_353h and Satt216 on MLG D1b (chromosome 2) in the HI population (Table 1).

DFRs Cosegregate with *W3* or *W4*

Glycine max DFR1 (AF167556; deposited by C. T. R. Tang, C. S. Wang, Taiwan Agricultural Research Institute, 1999), which contains the sequence part completely homologous to that of pDFR200 (Fasoula et al. 1995), except for 2 degenerate primer sequences, was used to amplify genomic DNAs from Hwangkeum and IT182932 via PCR. The genomic DNA sequences were employed to generate the SN016 marker. The genomic sequences of the *DFR2* gene

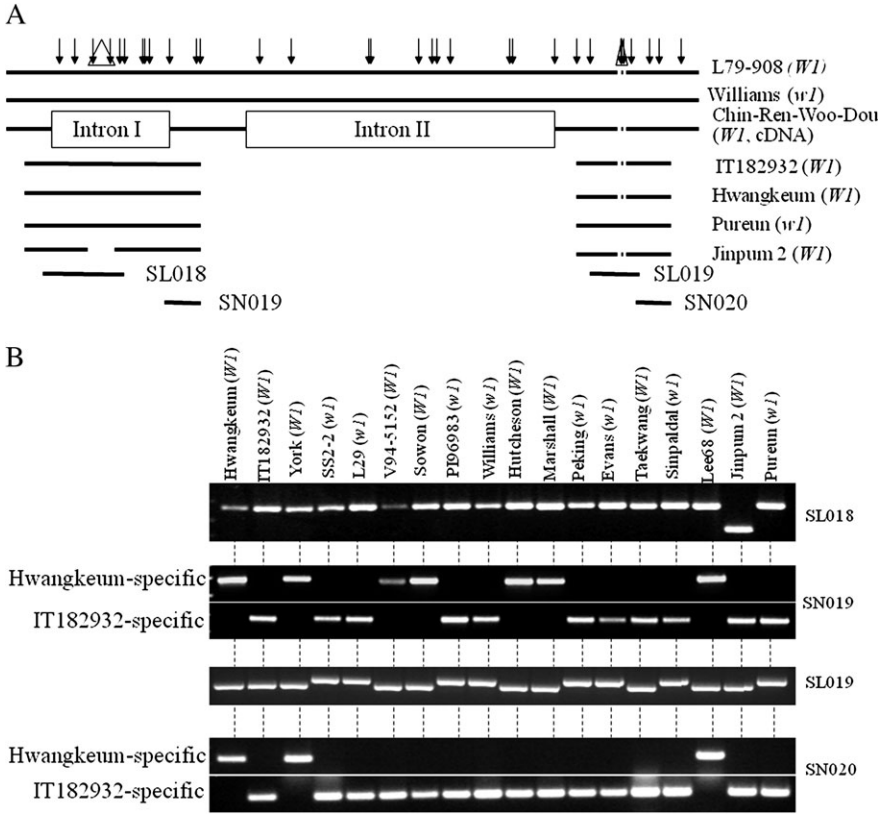

Figure 5. Schematic representation showing positions of the SL018, SN019, SL019, and SN020 markers on the genomic DNA segments and a full-length cDNA sequence encoding for *W1* alleles (Zabala and Vodkin 2007) (**A**) and agarose gels showing the separation of the alleles of the sequence-based markers SL018, SN019, SL019, and SN020 in 18 diverse soybean accessions with *W1* or *w1* genotypes (**B**). SNP sites are indicated by arrows and 2 large indel sites are indicated by triangles. The SN019 and SN020 markers were assayed using an allele-specific PCR procedure, as described by Jeong and Saghai Maroof (2004). The lanes that contain the common soybean accession between gels are joined by lines.

were identified as the *W4* locus through biochemical and molecular analyses of several *w4-m* lines arisen through insertion and reversion of a transposon (Xu et al. 2010). The 2 allelic genomic sequences of *DFR2* (EF187612 and DQ026299) were used to design primers for PCR amplification of their genomic DNAs from Hwangkeum and IT182932. The relevant genomic DNA sequences from both Hwangkeum and IT182932 were employed to generate the SL016 marker. The *DFR1* and *DFR2* cDNA sequences retrieved a third group of soybean ESTs (>85% identical) from a BLASTN search against GenBank est_others database. The EST contig, which was subsequently designated as *DFR3*, was used to design primers for PCR amplification of their genomic DNAs from Hwangkeum and IT182932. The genomic DNA sequences were utilized to generate the SM005 marker. The *DFR3* gene also was located on the completely sequenced *G. max* BAC clone gmw1-60n23 (AC170861; deposited by R. Shoemaker, I. Sanders, B. A. Roe, 2007). The gmw1-60n23 sequence region adjacent to the *DFR3* gene was scanned for microsatellite repeat presence and used to design a pair of primers. The resulting marker, SM325, cosegregated with SM005 from the EST contig of the *DFR3* gene, thereby confirming that DFR3 was located on gmw1-60n23.

Markers developed from the *DFR1*, *DFR2*, and *DFR3* genes were mapped on B2 (chromosome 14), D2 (chromosome 17), and D1b (chromosome 2), respectively (Table 1). Markers developed from the *DFR2* sequence were mapped between Satt386 and Sct_137 on MLG D2 (chromosome 17). This map location corresponds to the genetic location of the *W4* locus (Xu and Palmer 2005) and is also in agreement with the results that the *DFR2* gene encodes for the *W4* locus. At the same time, the *DFR2* sequence was mapped to the expected location in the Minsoy (*W4*) × T369 (*w4*) population used by Xu and Palmer (2005). *DFR1* was mapped between Sat_287 and Satt467 on MLG B2 (chromosome 14), and *DFR3* was mapped between Satt579 and Satt428 on MLG D1b (chromosome 2) (Table 1). The results of the Southern blot analysis suggested that *DFR* is encoded for by 2 or 3 genomic sequences and the DNA fragment that hybridized the most strongly with pDFR200 (*DFR1*) cosegregated with the *W3* locus (Wang et al. 1994;

Fasoula et al. 1995). Excluding the *DFR2*, which cosegregates with the *W4* locus, our results indicated that *DFR1* of the 3 *DFR* genes is most likely associated with the *W3* locus.

ANS (Leucoanthocyanidin Dioxygenase)

The soybean *ANS* gene sequence AF325853 (deposited to GenBank by J. M. H. Chiu and C-S. Wang in 2000) was found 85% identical to an *Arabidopsis ANS* Q96323 (Pelletier et al. 1997) in a BLASTX search against the GenBank nr database and retrieved 13 soybean ESTs from a BLASTN search against the GenBank est_others database. The alignment of the 13 ESTs clearly revealed 2 paralogous EST groups, although several SNPs that had probably originated from sequencing errors were observed. In order to PCR-amplify genomic DNAs from Hwangkeum and IT182932, specific primer sets were generated for each of the EST groups. The exon portions of the Hwangkeum genomic DNA sequence corresponded to one EST group were 99% similar to AF325853, and this genomic DNA sequence were referred to as the *ANS1* genomic DNA sequence. The genomic sequences of *ANS1* from Hwangkeum and IT182932 were compared and then were utilized in the generation of the SL015 and SN014 markers. As a result, *GmANS1* was mapped between BE806308 and Sat_272 on MLG B1 (chromosome 11) in the HI population (Table 1).

The exon portions of the Hwangkeum genomic DNA sequence that corresponded to the other EST group were 99% similar to part of *ANS2* (GenBank accession no. AY382829) and *ANS3* (AY382830), sequences which differed only at the 3′ and 5′ ends. Interestingly, at the G/T SNP site between Hwangkeum and IT182932 genomic DNA sequences, the T base appears only at the IT182932 sequence among Hwangkeum, IT182932, ANS2, and ANS3 sequences. At one other SNP C/G site, the C base appears at Hwangkeum and IT182932 and the G base at the GenBank *ANS2* and *ANS3* sequences. The results showed that, even if there were different 5′ ends and different 3′ polyadenylation sites, *ANS2*, *ANS3*, and the second EST group were mostly likely generated from the same gene. Thus, these genomic DNA sequences were designated as the *ANS2* genomic DNA sequence. A polymorphic locus that resides in a predicted intron of *ANS2* was utilized to develop the SN015 marker. The marker was mapped between Sat_414 and Satt129 on MLG D1a (chromosome 1) (Table 1).

MYB Transcription Factor

The soybean EST sequences, BM093788 and BM092559, >70% identical to the MYB transcription factor encoding for the tomato *ANT1* gene (Mathews et al. 2003) were exploited to generate the SL017 marker. The marker was mapped between Sat_009 and Satt070 on MLG B2 (chromosome 14) where the *W2* locus was mapped (Takahashi et al. 2008).

Anthocyanidin Reductase

A soybean EST (GenBank accession no. EV264997) with >86% sequence identity to the *Medicago truncatula* BANYULS gene (Xie et al. 2003) was exploited to amplify genomic DNA from Hwangkeum and IT182932 and then to generate the SN013 marker. SN013 was mapped between Satt207 and Satt493 on MLG A2 (chromosome 8) where the *O* locus was reported on the classical genetic map (Weiss 1970; use http://www.soybeanbreederstoolbox.org/index for the reference soybean classical map).

UPD-Flavonoid Glucosyltransferase

The results of genetic and biochemical studies indicated that the *R* gene is a structural or regulatory protein which is involved in the production of anthocyanins from leucoanthocyanidins (Buzzell et al. 1987; Todd and Vodkin 1993). Thus, the *R* gene may encode ANS, DFR, UFGT, or regulatory proteins. The *R* locus was mapped using the phenotypes of 40 RIL lines including black and buff seed coats, which were presumed to harbor the dominant R genotype, in the HI population (Supplementary Figure 1). The mapping results described above clearly indicated that the currently known soybean *ANS* and *DFR* genes are not linked to the *R* locus. The glycosyltransferase genes comprise a large multigene family. In the genome of *Arabidopsis*, more than 60 genes encoding for glycosyltransferases were annotated (The Arabidopsis Genome Initiative 2000). The sequence of the *Arabidopsis* UGT78D2 gene involved in the glucosylation of anthocyanidins (Lee et al. 2005) was used in a search for the soybean UFGT sequence from the GenBank database, via BLAST searches. As a result, groups of ESTs and BAC sequences were retrieved that were highly homologous with the UGT78D2 sequence (expect values $<10^{-15}$ for the translated protein sequence). Among those, 2 EST groups and 4 BAC sequences were mapped. However, none of them were mapped to MLG K (chromosome 9), wherein the *R* locus is located (Table 1).

Discussion

In this study, we have located 8 enzyme-encoding gene families and a transcription regulator involved in anthocyanin metabolism in a genome-wide soybean genetic map. This study allowed us to establish genetic correlation between the genes and all known soybean seed coat and flower color-controlling genetic loci, with the exception of the *R* locus.

Genetic Correlations between Seed Coat Color Traits and Genes Involved in the Anthocyanin Biosynthesis Pathway

Among the 5 genetic loci (*I*, *T*, *W1*, *O*, and *R*) shown to regulate seed coat pigmentation, the highly diverse markers among soybean lines for *I* and *T* correlated with previously known genetic map locations of *I* and *T* (Toda et al. 2002; Zabala and Vodkin 2003; Tuteja et al. 2004). The seed coat color genotypes of RIL in our mapping population and marker diversity in a set of soybean lines also supported genetic correlations between the *I* and *T* loci and the

markers. The genetic location of the genes encoding for *W1* was substantiated in the context of the current soybean molecular genetic map. A candidate gene, the soybean BANYULS gene, that may be associated with the *O* locus was suggested. The BANYULS genes from *Arabidopsis thaliana* and *M. truncatula* encode for anthocyanidin reductase, which converts anthocyanidins to their corresponding 2,3-cis-flavan-3-ols, thus promoting the accumulation of condensed tannins (Xie et al. 2003).

Genetic Correlations between Flower Color Traits and Genes Involved in the Anthocyanin Biosynthesis Pathway

Among the 6 genetic loci (*W1*, *W2*, *W3*, *W4*, *Wm*, and *Wp*) that regulate flower pigmentation, the genetic locations of the genes encoding for *W1*, *W4*, and *Wp* were substantiated in the context of the current soybean molecular genetic map. The *W3* locus cosegregates with the marker pDFR200 on the bases of RFLP analysis of isolines and an F_2 population segregating for the *W3* locus (Fasoula et al. 1995). By mapping a cDNA sequence containing the pDFR200 sequence, *W3* was localized, for the first time, to the soybean MLG B2 (chromosome 14). A candidate gene for *W2*, which is a MYB transcription factor, was suggested. The tomato *ANT1* gene encodes for an R2R3 MYB domain-containing transcription factor which regulates the genes involved in the both early and late steps of anthocyanidin biosynthesis, as well as the genes involved in anthocyanin modification and transport (Mathews et al. 2003). The map location of the soybean gene homologous to the *ANT1* gene corresponded to the genetic location of the *W2* locus (Takahashi et al. 2008). A regulatory gene, rather than a structural gene, in the anthocyanin biosynthesis pathway may encode for the *W2* locus (Takahashi et al. 2008). The *Wm* locus regulates levels of flavonol glycosides, not anthocyanins, and encodes *FLS* (Takahashi et al. 2007).

Genetic Correlations between Soybean Pigmentation Traits and Transcription Factors

This study demonstrated that, with the exception of *W2*, the genes encoding for enzymes involved in the anthocyanin biosynthesis pathway cosegregate with the genetic loci controlling soybean pigmentation. These results differ from the cases of maize, *Arabidopsis*, and *Solanaceae* plants, in that several transcription factors cosegregate with the pigment-controlling loci (De Jong et al. 2004; Lepiniec et al. 2006). The *R* gene was postulated to encode for ANS, UFGT, or regulatory proteins (Buzzell et al. 1987; Todd and Vodkin 1993). Thus far, none of the markers generated from these genes were mapped to the *R* locus located on MLG K (chromosome 9) (Song et al. 2004). Our scanning of the soybean whole genome shotgun sequence, Glyma1, released at the end of 2008 by Soybean Genome Project, DoE Joint Genome Institute (www.phytozome.net) indicated that no *ANS* or *UGT78D2* (*UFGT*) homologues are located in the vicinity of the *R* locus (data not shown), although the whole genome shotgun assembly might have a portion of misaligned genomic regions as shown in the rice genome sequencing efforts (International Rice Genome Sequencing Project 2005). Thus, the *R* locus is probably associated with a transcription factor, as more frequently shown in other plants.

Utility of the Developed Markers in Marker-Assisted Breeding and Association Studies

This study reports the degrees of correlations between polymorphisms in the anthocyanin biosynthesis genes and variation in seed coat and flower colors on a range of soybean accessions of known genotypes. The perfect correlations between seed coat color traits and genotypes of SM303, SM304, and SM305, which are 68, 64, and 14 kb, respectively, away from the inverted CHS gene repeats, are consistent with the previous observation that the *I* locus region has relatively extended levels of linkage disequilibrium, probably affected by a selective sweep during domestication (Hyten et al. 2007). This hypothesis should be tested more rigorously because the lack of correlation between seed coat color traits and genotypes of SM306, which is just 5 kb away from the inverted CHS gene repeats at the opposite side, needs to be explained. More diverse soybean accessions and additional markers need to be used. Polymorphisms of markers generated from sites presumed to induce functional defects at the *T* and *W1* loci always were associated completely with variation in color traits. In addition, this correlation study provides a glimpse of evolutionary relationships between the causative and non-causative polymorphic sites at the *W1* locus. The analysis of linkage disequilibrium in multiple soybean populations suggested that a whole-genome association analysis would require 9600–75 600 SNPs to identify successfully the most common haplotype variation (Hyten et al. 2007). Although limited, our correlation study suggested that more powerful makers, which will decrease the total number of SNPs, could be generated from sites presumed to induce functional defects and certain genomic regions, such as the *I* locus region, will require fewer numbers of SNPs. Thus, the results provide crucial empirical data for the design of future association studies in soybean.

General Conclusions

The soybean genome is generally regarded as an ancient diploidized tetraploid (Shoemaker et al. 1996). Soybean has undergone extensive rearrangements and additional duplications since its initial polyploidization (Lee et al. 2001). Thus, even in cases in which a gene encoding for a genetic locus was cloned in other plants, the mapping of a candidate gene cosegregating with a genetic locus in soybean (e.g., Lee et al. 2008) was not straightforward and required the development of 2 or more markers from highly homologous but unlinked gene family members, which are called homeologous gene families. In this study, 3 *DFR* homeologous genes and 2 *ANS* homeologous genes were mapped to different chromosomes. In contrast, *F3H1* and *F3H2*, highly similar to each other, cosegregated in our mapping population, which suggests that they are tandem repeats. Zabala and Vodkin (2005) have shown that F3H1

was the *Wp*-encoding gene on the basis of a transposon insertion event. *F3H2* is not likely responsible for the phenotype of the *Wp* locus.

The majority of soybean cultivars that are grown and consumed throughout the world today exhibit yellow (also called colorless or white) seed coats, whereas the majority of known accessions of the wild progenitor, *G. soja*, have black or, rarely, brown seed coats. Most of wild soybean seeds contain higher levels of natural products derived from the phenylpropanoid pathway, including proanthocyanidins, anthocyanins, and isoflavones, than is the case with cultivated soybean seeds. The markers presented in this report should facilitate the markers-assisted selection of seed coat colors in molecular breeding programs as the majority of color-controlling genes have been known to function epistatically. Additionally, the information will perform a crucial role in the further elucidation of the anthocyanin biosynthetic pathway in the soybean. Our strategy was similar to the study of identifying candidate genes for carotenoid accumulation loci by mapping all carotenoid biosynthetic enzyme genes in pepper (Thorup et al. 2000) and similar to that of identifying candidate genes for anthocyanin pigmentation loci by mapping 13 anthocyanin-related genes in the *Solanaceae* (De Jong et al. 2004). As De Jong et al. (2004) noted, further experimentation will be required to test the gene–locus correlation hypotheses suggested in the present study more rigorously.

Supplementary Material

Supplementary material can be found at http://www.jhered.oxfordjournals.org/.

Funding

BioGreen 21 Project (code no. 20080401034011 and 20050401034602), Rural Development Administration, the Republic of Korea, and, in part, by the Korea Research Institute of Bioscience and Biotechnology Research Initiative Program.

Acknowledgments

This is a joint publication from the Iowa Agriculture and Home Economics Experiment Station, Ames, IA, Project 4430 and from the USDA, Agricultural Research Service, Corn Insects and Crop Genetics Research Unit. The mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by Iowa State University or the USDA, and the use of the name by Iowa State University or the USDA implies no approval of the product to the exclusion of others that may also be suitable.

References

Benitez ER, Funatsuki H, Kaneko Y, Matsuzawa Y, Bang SW, Takahashi R. 2004. Soybean maturity gene effects on seed coat pigmentation and cracking in response to low temperatures. Crop Sci. 44:2038–2042.

Buss GR, Camper HM Jr, Roane CW. 1988. Registration of 'Hutcheson' soybean. Crop Sci. 28:1024–1025.

Buzzell RI, Buttey BR, MacTavish DC. 1987. Biochemical genetics of black pigmentation of soybean seed. J Hered. 78:53–54.

Cai CM, Van K, Kim MY, Jun T-H, Shin JH, Cho SY, Lee YS, Lee S-H. 2008. SNP discovery, linkage analysis and microsynteny in tentative consensus sequences derived from root cDNA in a supernodulating soybean mutant. Euphytica. 164:189–197.

Clough SJ, Tuteja JH, Li M, Marek LF, Shoemaker RC, Vodkin LO. 2004. Features of a 103-kb gene-rich region in soybean include an inverted perfect repeat cluster of CHS genes comprising the *I* locus. Genome. 47:819–831.

De Jong WS, Eannetta NT, De Jong DM, Bodis M. 2004. Candidate gene analysis of anthocyanin pigmentation loci in the *Solanaceae*. Theor Appl Genet. 108:423–432.

Dixon RA, Sumner LW. 2003. Legume natural products: understanding and manipulating complex pathways for human and animal health. Plant Physiol. 131:878–885.

Doyle JJ, Doyle JV. 1987. A rapid DNA isolation procedure for small amounts of leaf tissue. Phytochem Bull. 19:810–815.

Fasoula DA, Stephens PA, Nickell CD, Vodkin LO. 1995. Cosegregation of purple-throat flower color with dihydroflavonol reductase polymorphism in soybean. Crop Sci. 35:1028–1031.

Gore MA, Hayes AJ, Jeong SC, Yue YG, Buss GR, Saghai Maroof MA. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv*1 and *Rpv*1 region in soybean. Genome. 45:592–599.

Hegstad JM, Tarter JA, Vodkin LO, Nickell CD. 2000. Positioning the *wp* flower color locus on the soybean genome map. Crop Sci. 40:534–537.

Hyten DL, Choi I-Y, Song Q, Shoemaker RC, Nelson RL, Costa JM, Specht JE, Cregan PB. 2007. Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics. 175:1937–1944.

International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. Nature. 436:793–800.

Jeong SC, Saghai Maroof MA. 2004. Detection and genotyping of SNPs tightly linked to two disease resistance loci, *Rsv*1 and *Rsv*3, of soybean. Plant Breed. 123:305–310.

Lander ES, Green P, Abrahamson J, Barlow A, Daly JM, Lincoln SE, Newberg L. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics. 1:174–181.

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma T. 1993. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor Appl Genet. 86: 901–906.

Lee JH, Grant D, Vallejos CE, Shoemaker RC. 2001. Genome organization in dicots. II. Arabidopsis as a "bridging species" to resolve genome evolution events among legumes. Theor Appl Genet. 103:765–773.

Lee JH, Kang NS, Shin S-O, Shin S-H, Lim S-G, Suh D-Y, Baek I-Y, Park K-Y, Ha TJ. 2009. Characterisation of anthocyanins in the black soybean (*Glycine max* L.) by HPLC-DAD-ESI/MS analysis. Food Chem. 112:226–231.

Lee W-S, You J-A, Chung H, Lee Y-H, Baek N-I, Yoo J-S, Park YD. 2008. Molecular cloning and biochemical analysis of dihydroflavonol 4-reductase (DFR) from *Brassica rapa* ssp. *pekinesis* (Chinese cabbage) using a heterologous system. J Plant Biol. 51:42–47.

Lee Y, Yoon HR, Paik YS, Liu JR, Chung W-I, Choi G. 2005. Reciprocal regulation of *Arabidopsis* UGT78D2 and BANYULS is critical for regulation of the metabolic flux of anthocyanidins to condensed tannins in developing seed coats. J Plant Biol. 48:356–370.

Lepiniec L, Debeaujon I, Routaboul JM, Baudry A, Pourcel L, Nesi N, Caboche M. 2006. Genetics and biochemistry of seed flavonoids. Annu Rev Plant Biol. 57:405–430.

Mathews H, Clendennen SK, Caldwell CG, Liu XL, Connors K, Matheis N, Schuster DK, Menasco DJ, Wagoner W, Lightner J, et al. 2003. Activation tagging in tomato identifies a transcriptional regulator of anthocyanin biosynthesis, modification, and transport. Plant Cell. 15:1689–1703.

Palmer RG, Kilen TC. 1987. Qualitative genetics and cytogenetics. In: Wilcox JR, editor. Soybeans improvement, production, and uses. 2nd ed. Agron. Monogr. 16. Madison (WI): ASA, CSSA, and SSA. p. 135–209.

Palmer RG, Pfeiffer TW, Buss GR, Kilen TC. 2004. Qualitative genetics. In: Boerma HR, Specht JE, editors. Soybeans: improvement, production, and uses. 3rd ed. Madison (WI): ASA, CSSA, and SSSA. p. 137–214.

Pelletier MK, Murrell JR, Shirley BW. 1997. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in Arabidopsis. Further evidence for differential regulation of 'early' and 'late' genes. Plant Physiol. 113:1437–1445.

Rode MW, Bernard RL. 1975. Inheritance of a tan saddle mutant. Soybean Genet Newslett. 2:39–42.

Rozen S, Skaletsky HJ. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S, editors. Bioinformatics methods and protocols: methods in molecular biology. Totowa (NJ): Humana Press. p. 365–386.

Senda M, Masuta C, Ohnishi S, Goto K, Kasai A, Sano T, Hong JS, MacFarlane S. 2004. Patterning of virus-infected *Glycine max* seed coat is associated with suppression of endogenous silencing of chalcone synthase genes. Plant Cell. 16:807–818.

Shoemaker RC, Polzin K, Labate J, Specht J, Brummer EC, Olson T, Young N, Concibido V, Wilcox J, Tamulonis JP, et al. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics. 144: 329–338.

Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB. 2004. A new integrated genetic linkage map of the soybean. Theor Appl Genet. 109:122–128.

Stephens PA, Nickell CD. 1992. Inheritance of pink flower color in soybean. Crop Sci. 32:1131–1132.

Takahashi R. 1997. Association of soybean genes *I* and *T* with low-temperature induced deterioration. Crop Sci. 37:1755–1759.

Takahashi R, Asanuma S. 1996. Association of *T* gene with chilling tolerance in soybean. Crop Sci. 36:559–562.

Takahashi R, Matsumura H, Oyoo ME, Khan NA. 2008. Genetic and linkage analysis of purple–blue flower in soybean. J Hered. 99:593–597.

Takahashi R, Stephen M, Hatayama GK, Dubouzet EG, Shimada N, Aoki T, Ayabe S, Iwashina T, Toda K, Matsumura H. 2007. A single-base deletion in soybean flavonol synthase gene is associated with magenta flower color. Plant Mol Biol. 63:125–135.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature. 408:796–815.

Thorup TA, Tanyolac B, Livingstone KD, Popovsky S, Paran I, Jahn M. 2000. Candidate gene analysis of organ pigmentation loci in the *Solanaceae*. Proc Natl Acad Sci U S A. 97:11192–11197.

Toda K, Yang D, Yamanaka N, Watanabe S, Harada K, Takahashi R. 2002. A single-base deletion in soybean flavonoid 3′-hydroxylase gene is associated with gray pubescence color. Plant Mol Biol. 50:187–196.

Todd JJ, Vodkin LO. 1993. Pigmented soybean (*Glycine max*) seed coats accumulate proanthocyanidins during development. Plant Physiol. 102:663–670.

Todd JJ, Vodkin EO. 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. Plant Cell. 8:687–699.

Tuteja JH, Clough SJ, Chan WC, Vodkin LO. 2004. Tissue-specific gene silencing mediated by a naturally occurring chalcone synthase gene cluster in *Glycine max*. Plant Cell. 16:819–835.

Wang CS, Todd JJ, Vodkin LO. 1994. Chalcone synthase mRNA and activity are reduced in yellow soybean seed coat with dominant *I* alleles. Plant Physiol. 105:739–748.

Weiss MG. 1970. Genetic linkage in soybeans. Linkage group VII. Crop Sci. 10:627–629.

Woodworth CM. 1921. Inheritance of cotyledon, seed-coat, hilum, and pubescence colors in soy-beans. Genetics. 6:487–553.

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. 2003. Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science. 299:396–399.

Xu M, Brar HK, Grosic S, Palmer RG, Bhattacharyya MK. 2010. Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. Genetics. 184:53–63.

Xu M, Palmer RG. 2005. Genetic analysis and molecular mapping of a pale flower allele at the *W4* locus in soybean. Genome. 48:334–340.

Yang K, Moon J-K, Jeong N, Back K, Kim HM, Jeong SC. 2008. Genome structure in soybean revealed by a genomewide genetic map constructed from a single population. Genomics. 92:52–59.

Zabala G, Vodkin LO. 2003. Cloning of the pleiotropic *T* locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3′ hydroxylase. Genetics. 163:295–309.

Zabala G, Vodkin LO. 2005. The *wp* mutation of *Glycine max* carries a gene-fragment-rich transposon of the CACTA superfamily. Plant Cell. 17: 2619–2632.

Zabala G, Vodkin LO. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. Crop Sci. 47:S113–S124.

Received November 24, 2009; Revised May 7, 2010; Accepted May 30, 2010

Corresponding Editor: Prem Jauhar